M. Matveieva[1]
P. Polishchuk[1]
M.T.D. Cronin[2]

INTERPRETATION OF QSAR MODELS: THE IMPORTANCE OF MOLECULAR CONTEXT IN MINING STRUCTURAL PATTERNS

[1] Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University and University Hospital in Olomouc, Hněvotínská Str., 1333/5, Olomouc, Czech Republic;
[2] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK

mariia.matveieva01@upol.cz

Interpretation of QSAR models is essential to understand the nature of the biological effects of molecules as well as assisting in many practical aspects of drug development. Recently developed approaches can be applied to interpret QSARs regardless of the descriptors and/or machine learning methods used. One such approach calculates the contributions of any fragment to the property modelled [1, 2]. Those contributions may be averaged for identical fragments across the whole data set providing knowledge of whether fragments contribute "positively" or "negatively" to the activity. Such analysis reveal SAR trends captured by the model. There is, however, a drawback: this approach does not take into account the molecular environment. Molecular fragments should be assessed only with respect to their molecular context. The goal of this study was to develop a workflow to identify different molecular surroundings and their impact on a fragment's behaviour. We applied clustering based on a Gaussian mixture model which aimed to identify groups of compounds (clusters) comprising the same fragments where those fragments had a substantially different influence on the property modelled.

The approach was applied to analyse the toxicity of 1984 compounds to *Tetrahymena pyriformis*. Training set compounds were exhaustively fragmented based on the specific SMARTS pattern matching bonds to cleave. We analysed the distribution of contributions of fragments to the toxicity calculated from different molecules. These distributions can either have a bell shape (approximately normal), can be skewed, or have several peaks forming clusters potentially discerning the molecular context of corresponding fragments.

The results show that the clustering technique correctly identifies known toxicophoric patterns. In many cases it was able to distinguish molecular environments in which a fragment became potentially toxic. For example, when analysing halogens, we found that Cl, Br and I had consistently greater contributions to the toxicity when they were located in *alpha* position to a carbonyl group (e.g. α-haloketones or esters). For cyano groups, much greater contributions were detected in α-halonitriles. Methylcarbonyl groups conjugated with a double bond, e.g. in α,β-unsaturated ketones had substantially greater contributions. All these patterns can be linked to reactive mechanisms of toxic action as they represent fragments able to participate in nucleophilic substitution or addition reactions damaging cell membrane, proteins or other biomolecules. Thus, our findings are consistent with known data on toxicophores and mechanisms of action. This confirms the method has the potential to discovering patterns important for biological activities and interpretation of mechanisms of action.

1. Polishchuk P. et al. *Molecular Informatics*, 2013, **32:** 843–853.
2. Polishchuk P. et al. *Journal of Chemical Information and Modeling*, 2016**, 56:** 1455-1469.